EMPLOYMENT OF GEOGRAPHIC INFORMATION SYSTEMS FOR DETERMINING THE ACCURACY OF MUSEUM VOUCHER SPECIMEN DATA

KELLY E. ALLEN, ROBERT D. BRADLEY, R. RICHARD MONK, OLEKSIY V. KNYAZHNITSKIY,

NICK C. PARKER, DAVID J. SCHMIDLY, AND ROBERT J. BAKER

Voucher specimens in museums are the foundation of our understanding of systematics, biodiversity, evolutionary genetics, biogeography, ecology, etc. Today, many economic, political, and scientifically important decisions are made based on the synthesis of knowledge of biological organisms and associated biocomplexity. Therefore, the identification of voucher specimens and, subsequently, the conclusions drawn from knowledge based on these specimens is critical to many aspects of society.

Classically, identification of voucher specimens was verified and maintained through the efforts of specialists for a given taxonomic group or geographic region. Changes in nomenclature, taxonomic status, and geographic distribution or regional status were determined by the specialist and easily updated in the museum collection. The constant addition of specimens over time has produced museum collections of remarkable sizes. As these collections increased in size, curation of collections, maintenance of taxonomic order, and the ability to choose or develop museum computer databases have proven to be challenging and labor intensive (Sarasan, 1981; Sarasan and Neuner, 1983; Monk, 1996; Monk, 1997). The classical 'hands-on' approach has been taxed by the size of collections and the ever-decreasing number of qualified specialist in a given field or discipline. For example, a recent survey of mammal collections in the Western Hemisphere indicates 17% or approximately 58 collections consist of 10,000 or more specimens (Hafner et al., 1997). Additionally, at Texas Tech University a Recent mammal collection was initiated in the early 1960's, contained 5000 specimens in 1967, and today includes over 80,000 voucher specimens making the management of this collection a complex task.

The development of computers and computational technology since the 1960's has provided a means for better management of large sources of data. Increased memory, data storage, software, manipulation, retrieval, analysis capabilities, and archival quality printing capabilities have substantially improved since the initial development of the computer (DeMers, 2000). Collection of digital environmental data through aerial photographs, satellite imagery, global positioning systems (GPS), and geographic information systems (GIS) has provided a means to assess issues of biocomplexity that previously were difficult or impossible (Baker et al., 1998; Parker et al., 1998). In addition, recent technological advances not only provide a means to store and manipulate large data sets, but permit new insights into potential ecological and taxonomic variation for voucher specimens and thereby provide a means to identify possible problems associated with some voucher specimens.

At Texas Tech University, comparison of documented distributions of Texas mammal species (Davis and Schmidly, 1994) with maps representing distributions created from habitat associations and known collection localities (Allen, 2000) resulted in the identification of several voucher specimens that were either possibly misidentified, represented significant range extensions (Figs. 1a, 1b, 1c), or required nomenclature changes. Several possibilities could result in the observed mismatches of the known collection localities. A full summary of these possibilities is given by Allen (2000), but can be narrowed to (1) the expansion of a species' range extent, (2) a limitation of the digital data resolution or scale, (3) temporal changes in habitat, (4) a nomenclature change, and (5) the misidentification of specimens. In this paper, the misidentification of voucher specimens refers to the assignment of an incorrect taxonomic name for a voucher specimen within the museum archive. This may be the product of misidentification of the voucher specimen or an error due to data handling (Monk, 1997). In this study, we examined voucher specimens for which there appeared to be a mismatch with either the documented range, predicted habitat, or actual specimen locality to explore which possibilities accounted for the apparent inconstancies observed in the data sets.

METHODS

Geographic coordinates were assigned to all mammalian voucher specimens collected in Texas and housed within the Natural Science Research Laboratory (NSRL), Museum of Texas Tech University. Coordinates were assigned using a computer algorithm that determines the Universal Transverse Mercator (UTM) coordinate based upon the traditional descriptor location (Knyazhnitskiy et al., 2000). Assignment of UTM coordinates to voucher specimen localities provided a geographic location for that voucher specimen that could be easily translated by a GIS.

Maps illustrating predicted habitat for 127 native, terrestrial mammals were acquired from the Texas GAP Analysis Program. Predictive habitat maps were generated with the use of a database of known habitat associations for each mammalian species (Figs. 1.b and 1.c). A GIS was used to query predictive habitat from digital data sets consisting of vegetation, soils, hydrography, and documented range extents (Fig. 1.a).

Predicted habitat for each mammal was modeled using two methods. The first method modeled associated habitat throughout the state and reflects possible breeding habitat (Fig. 1.b). The second model indicates predicted habitat restricted within the documented range extent (Fig. 1.c).

Once the voucher specimen data were geocoded, they were input into a GIS and combined with digital data consisting of documented range and predicted habitat maps for 127 species of mammals from Texas. A map of each species' range, predicted habitat, and respective voucher specimens was produced and visually examined for apparent mismatches (Figs. 2.a, 2.b, 2.c). When localities did not match either the documented range extent or the predicted habitat, the voucher specimen was examined by mammalogists for specific identification, or was identified as a nomenclature change resulting in modification begin made to the digital data set using a GIS.

RESULTS AND DISCUSSION

Approximately 30,000 specimens were plotted in association with their respective species' documented range extent and predicted habitat. Of these, 1073 (3.6%) voucher specimen records did not correspond with their respective documented range or predicted habitat. As a result, 221 voucher specimens were examined by mammalogists and 290 specimen records were examined using a GIS. In addition, 527

records of *Perognathus flavus* were recognized as mismatches with the documented range extent. However, these could not be verified at this time due to the complicated nature of the *Perognathus* complex. An additional 35 specimens were unaccounted for because either they were missing, had been deaccessioned, or were not pulled because it was obvious the mismatch resulted from a database error. After verification, 332
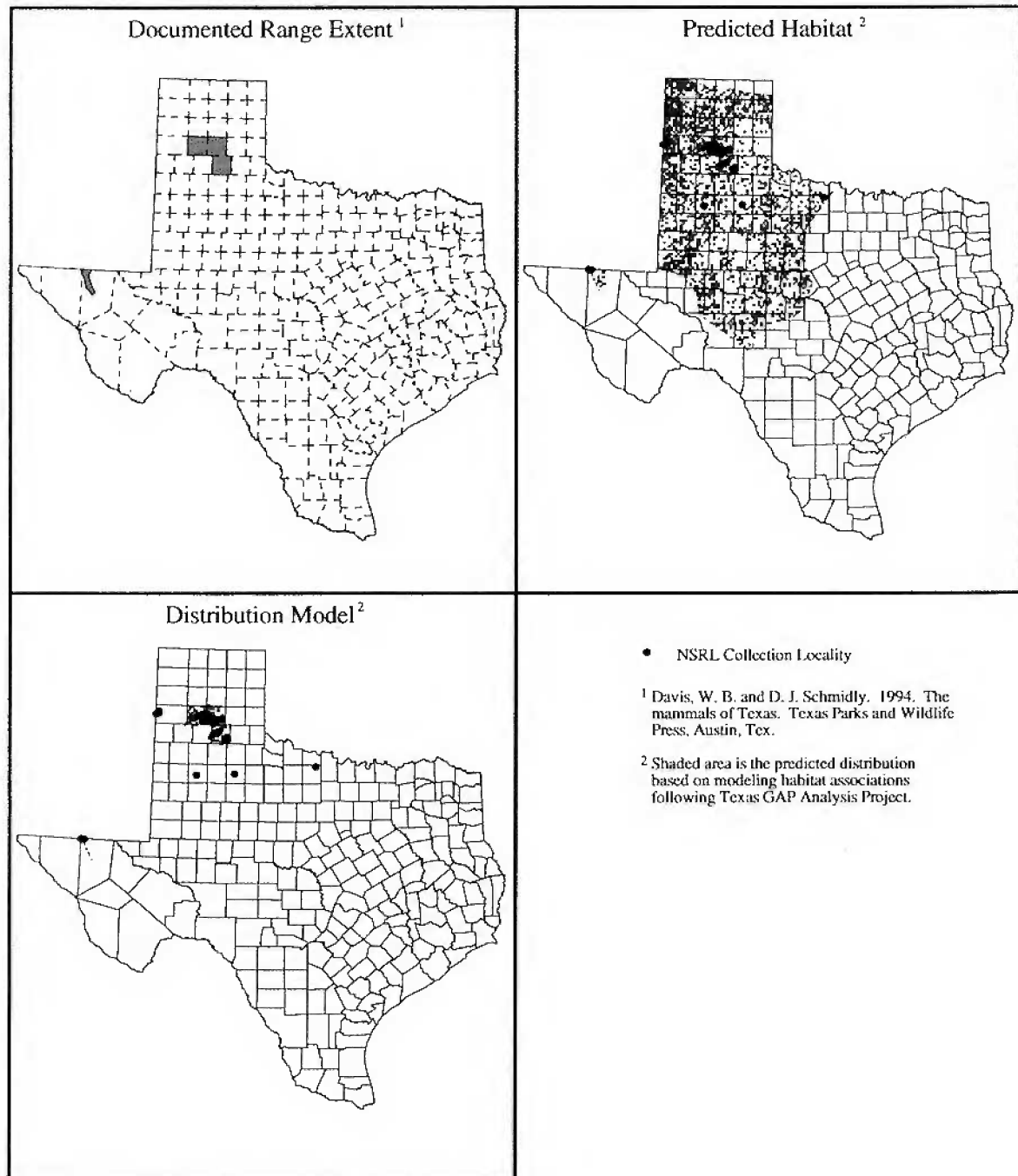
PINON MOUSE
Peromyscus truei
AMAFF03130

Documented Range Extent[1]

Predicted Habitat[2]

Distribution Model[2]

• NSRL Collection Locality

[1] Davis, W. B. and D. J. Schmidly. 1994. The mammals of Texas. Texas Parks and Wildlife Press, Austin, Tex.

[2] Shaded area is the predicted distribution based on modeling habitat associations following Texas GAP Analysis Project.

Figure 1. Predicted habitat models for *Peromyscus truei* model. Models were created using a Geographic Information System and reflect (a) the documented range extent as published in Davis and Schmidly (1994), (b) predicted habitat occuring throughout the state and (c) predicted habitat restricted by the documented range extents.

A. *Peromyscus truei*

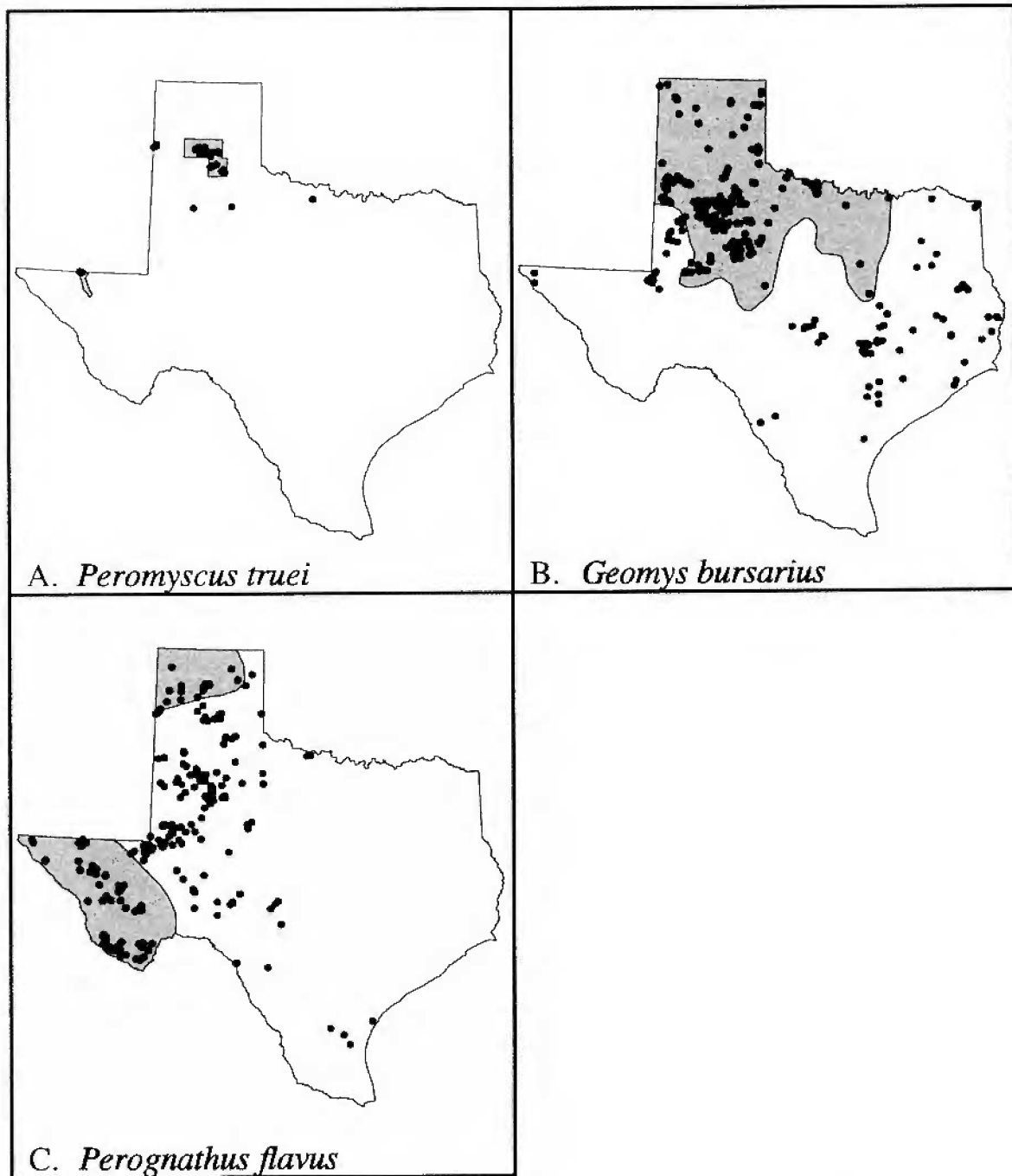B. *Geomys bursarius*

C. *Perognathus flavus*

Figure 2. Illustration of mismatches observed between voucher specimen localities and their representative documented range extents for (a) *Peromyscus truei*, (b) *Geomys bursarius*, and (c) *Perognathus flavus*. Shaded regions represent the documented range extents digitized from Davis and Schmidly (1994).

(31%) of the voucher specimen records initially identified as possible mismatches and comprising 20 species, 10 genera, and 5 families were found to be incorrectly identified to species or required nomenclature changes. Of these 332 specimens records, 313 (94%) required changes in nomenclature whereas 20 (6%) were the result of misassignment upon cataloguing. Examples of voucher specimens misidentified as a result of changes in nomenclature include *Dipodomys ordii*, *Peromyscus attwateri*, *Peromyscus difficilis*, *Peromyscus boylii*, and *Peromyscus nasutus*, and the *Geomys* complex. Examples of voucher specimens misidentified upon initial cataloguing include *Tayassu tajacu*, *Myotis volans*, *Lasiurus ega*, and *Reithrodontomys fulvescens*.

Two special cases of extensive mismatch of specific assignment of voucher specimens in the museum database and the ecological/geographic range involved the *Geomys bursarius* complex (Fig. 1.b) and the *Perognathus flavus* complex (Fig. 1.c). Both of these involved a substantial number of voucher specimens, *Geomys* complex, 290 records, and *P. flavus*, 527 records. The problem associated with *Geomys* resulted from specimens previously recognized as a single morphologically highly variable species (Hall, 1981) being subdivided into six different species (*G. attwateri*, *G. breviceps*, *G. bursarius*, *G. knoxjonesi*, *G. personatus*, and *G. texensis*) (Baker et al., 1989; Block and Zimmerman, 1991; Bradley, et al., 1991; Honeycutt and Schmidly, 1979; Tucker and Schmidly, 1981). Taxa within the genus *Geomys* exhibit allopatric distribution. Due to a lack of concensus regarding taxonomic arrangements, the museum database had not been updated. In addition to the specimens of *G. bursarius* being misidentified due to nomenclature changes, several specimens were found to be misidentified based upon the geographic location of the collection locality. Revisions of the museum database were based on the published ranges recognized in the above-cited papers and in The Mammals of Texas (Davis and Schmidly, 1994).

The problem with the *Perognathus flavus* complex is more difficult because this species complex involves two sibling (cryptic) species (*P. flavus* and *P. merriami*) that are often difficult to distinguish based on traditional skin and skull characteristics (Lee and Engstrom, 1991). Based upon karyology and biochemi-

cal methods, *P. flavus* and *P. merriami* have been documented to be reproductively isolated but karyological and biochemical data are not available for all museum specimens. The proper assignment of voucher specimens must await more detailed studies of the museum specimens and the development of morphological methods of species identification. At present, it seems best to recognize that in the *P. flavus/merriami* complex "specific" assignments in this complex are tenuous unless biochemical or chromosome data are available and that ecological and other conclusions based on these specimens will need to be viewed in light of this situation.

The methods employed in this study assisted the curators by increasing accuracy of the museum database. The methods are limited, however, in that they only identify ecological or geographical outliers. Specimens from the assigned geographic and ecologic ranges also may be misidentified. These misidentifications cannot be recognized using the methodology described in this paper because the voucher specimens appear to occur within the appropriate geographic or ecological range. Nonetheless, the methods employed here can be added to those available to museum curators to strengthen the accuracy of museum databases.

In addition, the authors believe this method provides a rapid, cost-effective means for the non-specialist to identify potential problems. A non-specialist here refers to someone that is not capable of verifying the vouchers species for determining the accuracy of a museum's database. Obviously this method requires museums to have already converted descriptor locations of each voucher specimen to some format that is acceptable for a GIS. The conversion of a descriptor locality to one that can be interpreted by a GIS can indeed be very costly and time consuming to a museum. In this situation, the non-specialist would have to assume the museum's specimens are correctly identified and reflect current nomenclature status. However, if the conversion to digital data already has been made and the location of voucher specimens is known by the museum, the comparison of documented range extents with voucher locations is easily done using an algorithm that automates the process of mapping the documented range extent with the geographic location of voucher specimens. Visual identification of mismatches can then be completed in a matter of hours.

The authors would like to stress that the application of GIS is no substitute for a knowledgeable specialist, a workable key, and an accurate range map for the purpose of identifying specimens. The strength of any set of data is a function of the collector and the means of processing the data. Any identified mismatches using the methods outlined in this paper should be further addressed by a taxonomic specialist.

Literature Cited

Allen, K. E. 2000. Areographic Fragmentation of Texas Mammal Distributions: A Fractal Analysis. Ph.D. Dissertation, Texas Tech Univ., vii+1-261pp.

Baker, R. J., S. K. Davis, R. D. Bradley, M. J. Hamilton, and R. A. Van Den Bussche. 1989. Ribosomal-DNA, Mitochondrial-DNA, Chromosomal and Allozymic Studies on a Contact Zone in the Pocket Gopher, *Geomys*. Evolution, 43:63-75.

Baker, R. J., C. J. Phillips, R. D. Bradley, J. M. Burns, D. Cook, G. F. Edson, D. R. Haragan, C. Jones, R. R. Monk, J. T. Montford, D. J. Schmidly, and N. C. Parker. 1998. Bioinformatics, Museums and Society: Integrating Biological Data for Knowledge-based Decisions. Occas. Pap., Mus., Texas Tech Univ. 187:i+1-4.

Block, S. B. and E. G. Zimmerman. 1991. Allozymic Variation and Systematics of Plains Pocket Gophers (*Geomys*) of South-Central Texas. The Southwestern Naturalist, 36:29-36

Bradley, R. D., S. K. Davis, and R. J. Baker. 1991. Genetic Control of Premating-Isolating Behavior: Kaneshiro's Hypothesis and Asymmetrical Sexual Selection in Pocket Gophers. J. Heredity, 82:192-196.

Davis, W. B. and D. J. Schmidly. 1994. The Mammals of Texas. Austin: Texas Parks and Wildlife.1-338pp.

DeMers, M. N. 2000. Fundamentals of Geographic Information Systems. 2nd Ed. New York: John Wiley & Sons. iii+1-498.

Hall, E. R. 1981. The Mammals of North America. 2nd Ed., New York: John Wiley and Sons. Vol. 1:1-600; vol. 2:601-1181.

Honeycutt, R. L. and D. J. Schmidly. 1979. Chromosomal and Morphological Variation in the Plains Pocket Gopher, *Geomys bursarius*, in Texas and Adjacent States. Occas. Pap., Mus. of Texas Tech Univ., 58:1-54.

Lee, T. E., Jr., and M. D. Engstrom. 1991. Genetic Variation in the Silky Pocket Mouse (*Perognathus flavus*) in Texas and New Mexico. J. Mamm., 72:273-285.

Knyazhnitskiy, O.V., R. R. Monk, N. C. Parker, and R. J. Baker. 2000. Assignment of Global Information System Coordinates to Classical Museum Localities for Relational Database Analyses. Occas. Pap., Mus. of Texas Tech Univ., 199:1-15.

Monk, R. R. 1996. An Automated Data Management System for use in the NSRL, Museum of Texas Tech University. Ph.D. Dissertation Proposal, Texas Tech Univ., 21 pp.

Monk, R. R. 1997. Automated Data Management in Systematics Collections. Ph.D. Dissertation, Texas Tech University, vii+118 pp.

Parker, N. C., R. D. Bradley, J. M. Burns, G. F. Edson, D. R. Haragan, C. Jones, R. R. Monk, J. T. Montford, C. J. Phillips, D. J. Schmidly, and R. J. Baker. 1998. Bioinformatics: A Multidisciplinary Approach for the Life Science. Occas. Pap., Mus. of Texas Tech Univ., 186:1-8.

Sarasan, L. 1981. Why Museum Computer Projects Fail. Museum News, Jan/Feb 1981:40-49.

Sarasan, L. and A. M. Neuner, 1983. Museum Collections and Computers. Association of Systematics Collections, 292 pp.

Tucker, P. K. and D. J. Schmidly. 1981. Studies on a Contact Zone among Three Chromosomal Races of *Geomys bursarius* in East Texas. J. Mamm., 62:258-272.

Addresses of authors:

KELLY E. ALLEN

Texas Cooperative Fish & Wildlife
Research Unit
Texas Tech University
Lubbock, TX 79413-2120

Current Address:
Central Oregon Community College
2600 NW College Way
Bend, OR 97701
e-mail: kallen@cocc.edu

ROBERT J. BRADLEY

Department of Biological Sciences and
Museum of Texas Tech University
Lubbock, TX 79409-3131
e-mail: rbradley@ttacs.ttu.edu

NICK C. PARKER

Texas Cooperative Fish & Wildlife Research Unit
U.S. Geological Survey-Biological Resources
Division, Texas Tech University
Lubbock, TX 79413-2120
e-mail: nparker@ttu.edu

OLEKSIY V. KNYAZHNITSKIY

Department of Biological Sciences and
Museum of Texas Tech University
Lubbock, TX 79409-3191
e-mail: alexey_khn@hotmail.com

DAVID J. SCHMIDLY

President, Texas Tech University
Department of Biological Sciences and
Museum of Texas Tech University
Lubbock, TX 79409-3131
e-mail: schmidly@ttu.edu

R. RICHARD MONK

Museum of Texas Tech University
Box 43191
Lubbock, TX 79409-3191
e-mail: rich.monk@ttu.edu

ROBERT J. BAKER

Department of Biological Sciences and
Museum of Texas Tech University
Lubbock, TX 79409-3131
e-mail: rjbaker@ttu.edu